# CSE 184 Final Project Design Document

### Job Posting Gender-Specific Pronouns Analysis

#### Kyle O’Brien, Bryan Jimenez, Nikhil Dodd, Jerico Factor

## Motivation

It's common knowledge that a majority of software engineers identify as male. Despite gains in gender diversity across other STEM professions, software engineering remains largely stagnant. There is a myriad of factors influencing this reality. I'm interested in exploring the role the wording of job postings play in potentially discouraging non-male candidates from applying. This project is inspired by [**Textio**](https://textio.com/?utm_source=bing&utm_medium=ppc&utm_campaign=bing-textio-flow&utm_content=ad1&utm_term=Textio&utm_campaign=google-competition&utm_source=bing&utm_medium=ppc&hsa_tgt=kwd-73804879382921:loc-190&hsa_grp=1180876345744712&hsa_src=o&hsa_net=adwords&hsa_mt=e&hsa_ver=3&hsa_ad=%7bcreative%7d&hsa_acc=2567025180&hsa_kw=Textio&hsa_cam=1415385531)**.**

**White Paper:** [Evidence That Gendered Wording in Job Advertisements Exists and Sustains Gender Inequality - Journal of Personality and Social Psychology](http://www.fortefoundation.org/site/DocServer/gendered_wording_JPSP.pdf?docID=16121)

[**Github Repo**](https://github.com/Kyle1668/gender-specific-pronouns-analysis)

**Questions**

1. Out of X randomly collected tech jobs, what percentage have gender-specific pronouns?
2. What’s the percentage breakdown by state/region do jobs have gendered phrases?
3. What are the most common keywords that encourage/discourage applicants?
4. Does the company size or location influence a company's inclusion initiatives?
5. What are the ten most common masculine phrases/wording used in tech job postings?

**Revised Questions - 11/14/2019**

1. What percentage have gender pronouns in the job posting?
2. How are gendered job postings distributed across the USA?
3. What are the most common roles in tech that have gendered job postings?
4. Out of the gendered words/phrases, which are the most common?
5. Find the correlation between phrase usage and location (are certain phrases being used more based on the origin of the company/where posting is coming from)

**Whos Interested**

Increasingly, the tech industry is focusing its efforts on addressing systematic ……

**Public Policy / Topics**

Inclusion in the tech industry, culture, and recruiting.

## Data

**Schema**

|  |  |  |
| --- | --- | --- |
| **Field** | **Data Type** | **Required** |
| Job Title | string | true |
| Company | string | true |
| Location | string | true |
| Date Posted | string | true |
| Applicants | integer | false |
| Views | integer | false |
| Seniority Level | string | true |
| Industry | string | true |
| Employment Type | string | true |
| Job Functions | string | true |
| Posting Body | string | true |

**Data Sets**

1. Created our own curated dataset from LinkedIn job postings. This dataset will have the above schema and be curated using web scraping.
2. The Kaggle [U.S. Technology Jobs on Dice.com](https://www.kaggle.com/PromptCloudHQ/us-technology-jobs-on-dicecom/activity) dataset from Kaggle. This dataset containers 22,000 tech industry job postings.

## Topical Articles

* [Here Are The Words That May Keep Women From ... - HuffPost](https://www.huffpost.com/entry/textio-unitive-bias-software_n_7493624)
* [Watch your (gender) tone How Textio reveals the unconscious bias that’s lurking in your job postings - Textio](https://textio.com/blog/watch-your-gender-tone/13035166463)
* [Part 1: 70 Inclusive Language Principles That Will Make You A More …](https://medium.com/diversity-together/70-inclusive-language-principles-that-will-make-you-a-more-successful-recruiter-part-1-79b7342a0923)
* [Part 2: 70 Inclusive Language Principles That Will Make You A ...](https://medium.com/diversity-together/70-inclusive-language-principles-that-will-make-you-a-more-successful-recruiter-part-2-eb0f5e4f6e80)
* <https://speakspeak.com/resources/english-grammar-rules/pronouns/personal-pronouns>
* <https://www.glassdoor.com/employers/blog/10-ways-remove-gender-bias-job-listings/>
* <http://gap.hks.harvard.edu/evidence-gendered-wording-job-advertisements-exists-and-sustains-gender-inequality>
* <https://www.businessnewsdaily.com/9434-gender-neutral-job-ad.html>

### 11/18/19 TA Meeting Notes

Instead of a Jupyter notebook, we should put our focus on an interactive web application.

<https://d3js.org/> -- Interactive JS for responsive web apps

We should get multiple data sets. Make ETL pipeline

**Tasks**

* Build ETL pipeline with LinkedIn REST API
* Build visualization using mock data
* Run analysis on data to answer questions
  + Nationally
  + By State
* Refactor mock data using live data
* Have suggested a news article

**Stretch Goals**

* **.**